

Chapter 1

Random Variables

1.1 Basic Concepts

In this chapter, we concern ourselves with the basic concepts involved in dealing with random numbers. The first question to be dealt with is, “What is a random number?” I will give plausible answers to this question rather than a formal definition. This chapter will deal with the formal calculus of statistics. Chapter 2 will use the concepts developed to do useful calculations.

A set of random numbers is a set of numbers for which knowledge of any subset of the numbers will not tell you *with certainty* the value of the next one sampled from the set. For certain sets of numbers like the outcome of the roll of a die, the possibilities are limited and you may occasionally actually guess the result of the next throw. The point is that you cannot do it with certainty. Otherwise, playing dice would not be gambling!

A typical situation in science or engineering is that you get a set of measurements from a system. The next measurement is not certain, no matter how many measurements you took before and no matter how accurately you took those numbers. This situation is typified by a process where you put a set of small pieces of paper in a box, each containing a different number. If the contents of the box are mixed up, you should not be able to predict the next number drawn by looking at the previous number drawn. If you know that no number is repeated, you are in the position of exactly knowing what numbers will *not* be drawn, but you cannot be certain what the next number will be.

In engineering and science, you are often in the position of having a system where the inputs are constant, but the output(s) fluctuate in a random manner. For instance, you attempt to measure the signal strength

from a cell phone 5 km away. The measured signal strength will vary with time because of the events occurring in the atmosphere between you and the cell phone. You will not be able to predict exactly how those fluctuations will occur. The field strength measurements are interfered with by the fluctuations. On the other hand, if you hook up a flow system to send water at a constant rate through a pipe, you will find that under certain conditions, everything you can control is constant, but the velocity of the flow at a given point is fluctuating. This flow condition is called turbulence and is often the subject of study itself.

The examples in the previous two paragraphs are typical in that the process that connects the inputs to the outputs is where the randomness is generated.

I have seen definitions that describe a set of random numbers as a set of numbers that takes at least as many bits to describe as is contained in the set. The easiest way to understand this concept is to imagine that the set of random numbers is all decoded into bits (*viz.* $6 = 0\ 1\ 1\ 0$) and then strung all together to make one big binary word. If it were possible to find repeating sets of bits in the binary word, then we could describe the set of numbers with fewer bits than appear in the original set. Let's make this more concrete by a simple example. Consider a set of coin flips of an "honest" coin. We will call "heads" 1, and "tails" 0. Make a 100 bit binary word by flipping the coin 100 times and recording the result of the n th trial in the n th place in a 100 bit binary word. If the coin were truly "honest", you would need all 100 bits to be guaranteed that you could uniquely describe the results of all your trials. Modern data compression schemes such as used in MP3 depend on the accidental redundancy in any particular realization of a set of 100 coin flips, but this does not change the need for all 100 bits to be guaranteed a unique representation.

In many instances, random numbers selected sequentially are not related to each other in the sense that if one particular number is larger than the mean, the next number could be higher or lower than the mean. In such a situation, it doesn't matter in what order you deal with the numbers. For instance, the errors in the adjacent measurements of the temperature on the rod mentioned above are usually not related to each other.

On the other hand, there are situations where the random numbers are related to each other. If a number is larger than the mean, the numbers next to it are likely to also be larger than the mean. Here, it matters what order the numbers are sampled in. This leads us to the concept of *the time series*. A time series is a set of random numbers where the order matters.

It is important to know which one is first, second, etc. This condition can arise when measuring the fluctuations in the field strength of a cell phone, if the time scale is small enough. Events in the atmosphere that cause the fluctuations can only change but so fast.

The result of a calculation involving random numbers is called a *statistic*. For instance, if you add together four samples of a set of random numbers, the resulting statistic would also be a random number. The study of the results of computations using random numbers is called *statistics*. A statistic well known to baseball fans in the US is the batting average. It is computed by dividing the number of hits that result in the player getting on base by the number of times the player attempts to bat times 1000. It is taken as a measure of the player's ability. Since no player has ever hit safely every time at bat through a season, the chances of the player hitting safely is assumed to be a random number proportional to his or her ability.

In studying statistics, you will have to define some strange, idealized objects that are never seen in real life. For instance, there is an implicit assumption made when doing the calculus of statistics that the numbers you are dealing with come from a common source called a *population*. In principle, there are an infinite number of elements in the population, but you can only deal with a sample of them. For instance, if you measure the diameter of axles made by some machine, the set of measurements made up from sample diameters of the machine output would be a sample of the population of all possible axles this particular machine could make. Implicit in this assumption is that the machine could have made an infinite number of samples without wear and tear in the cutting tools etc. You are dealing with the infinite variety of axles the machine could make in its current configuration. The settings of the tool define the state that creates the population.

If you have two machines making axles and you keep the measurements of the diameters separate, then you would be dealing with two populations. If, on the other hand, you just throw axles from both machines in a pile and make measurements later, without noting which machine the sample came from, you are back to dealing with a single population.

Suppose you are put in charge of making sure that all the axles shipped to a customer are within a certain diameter range. You would then measure every axle made by the machine. Suppose, also, that you get one axle that is outside the allowed range when all the previous ones had been within range. How do you tell if it is "just a fluke" or if you have evidence that

the machine's condition has changed and it is behaving in a substandard manner? Or, can you tell the difference between the two alternatives presented above?

Further, suppose you are put in charge of measuring the heat transfer coefficient for a new design for a compact heat sink designed by your company. If you measure the heat transfer coefficient for the same sample more than once, I guarantee that each measured heat transfer coefficient will be different. How close do you think you got with any of the measurements to the "right" value? How much variation in the results between experiments is reasonable?

Answers to the above and other related problems are derived through a formal calculus of statistics. This calculus involves manipulation of the formal objects hinted at above such as the probability distribution functions, the expected value, etc.

We start by examining the idealization of the average of a sample of random numbers. In statistics texts, what we informally call the average of something, is an approximation to the formal concept of the *expected value*.

1.2 Expected Values

If x is the outcome of some experiment with a random component, the practical definition of the expected value, $\langle x \rangle$, is as the average of an *infinite* number of trials of the experiment. This type of averaging is performed by doing the same experiment many times. It is called *ensemble averaging*. A more mathematical definition will be given below.

1.3 Probability Distribution Functions

Suppose an experiment could only have a finite number of outcomes. An example is the rolling of unloaded dice. Let the n th outcome of the rolling of a die be denoted by x_n . The average outcome can be computed by adding up the values of all the events and then dividing by the total number of events. Alternatively, since there are only a finite number of possibilities for the outcome, the same calculation could be performed by measuring the number of times each value of x_n occurs. Let $P(x_n)$ be the number of times a particular value of x_n occurred. A typical result from rolling a 6 sided die 20 times is the sequence

$$(3, 1, 4, 6, 6, 6, 1, 3, 6, 1, 2, 1, 1, 1, 2, 1, 2, 3, 4, 3)$$

Here, $P(1) = 7$, $P(2) = 3$, $P(3) = 4$, $P(4) = 2$, $P(5) = 0$, $P(6) = 4$.

The average of this sequence can be computed directly by adding up the numbers in the sequence and dividing by 20. The result is 2.85. You can also compute the average by grouping like results as we did in defining $P(\cdot)$. The sum of the results can be computed by adding up the products $x_n P(x_n)$. The number of results is $\sum_n P(x_n)$. The average can then be computed as

$$\bar{x} = \frac{\sum_n x_n P(x_n)}{\sum_n P(x_n)}. \quad (1.3.1)$$

In the example sequence above, the average is computed

$$\bar{x} = \frac{1 \times 7 + 2 \times 3 + 3 \times 4 + 4 \times 2 + 5 \times 0 + 6 \times 4}{7 + 3 + 4 + 2 + 0 + 4} = \frac{57}{20} = 2.85.$$

For a large number of rolls of a single die, a typical set of measurements would give a $P(\cdot)$ like that shown in Figure 1.1.

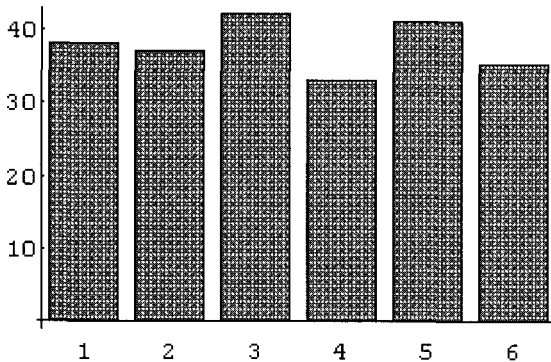


Fig. 1.1 Results of experiment rolling one die.

Experience has shown that if the experiment (rolling the die) were repeated a very large number of times, the ratio $P(x_n) / \sum P(x'_n)$ would tend to $1/6$. This ratio is called the *probability of the result* x_n . It should be clear that the probability is bounded by 0 and 1. If the result never occurs (probability is 0), we don't care about it. If it occurs all the time, (probability is 1), the process is not random.

Hence, in this case, $\langle x \rangle$, the infinite average, can be computed by using the limit of the weighting, assuming an infinite number of samples.

$$\langle x \rangle = \lim_{N \rightarrow \infty} \frac{\sum P(x)x}{\sum P(x')} = \frac{1}{6} \sum_{x=1}^6 x = \frac{21}{6} = 3.5. \quad (1.3.2)$$

The *expected* value is 3.5; however, note that the value is not the most likely value! In fact, you can't roll a die and get a value of 3.5. In this case, the expected value should be viewed as being the limiting value of the average of the outcomes, not as a statement about what outcomes are expected from each trial. For the example given here, the expected value is not a particularly useful concept — at least it gives incomplete knowledge about the data set. Please hang in there. More useful concepts will be given shortly.

The limit of the function

$$P(x_n) / \sum_k P(x_k) \quad (1.3.3)$$

as the number of samples goes to infinity is called the *probability distribution function* (pdf) for the event x_n . This function is denoted $p(x_n)$. It can be seen to be the expected relative frequency of event x_n in any sampling of the output. In this case, it can be computed *a priori* by assuming that the probability of an outcome is equal to the number of ways the outcome can occur, divided by the total number of possible outcomes. For this example, each outcome could only occur in one way and there were 6 equally possible outcomes. The *a priori* probability of any outcome of rolling a die is thus $1/6$. If the pdf is determined by measurement, it is called the *a posteriori* pdf.

The mathematical definition of the expected value for a random variable with a discrete output should now be apparent. It is defined in terms of the probability density function, *viz.*

$$\langle x \rangle = \sum_{\text{all possible } x} xp(x).$$

I will deal with random variable with a continuous output below.

If the function $p(x)$ had a peak in it, the value of $p(x)$ at the peak would be *the most likely outcome*.

Consider a more complicated experiment, rolling two dice at the same time and adding the result. Here the possible sums range from 2 to 12.

There are $6 \times 6 = 36$ possible outcomes.

There is only one way to get 2 or 12, $1 + 1$ or $6 + 6$, so the *a priori* probability of each of these is $1/36$. There are more ways to get 3, $1 + 2$, $2 + 1$. The *a priori* probability of 3 and 11 is $1/18$. Continuing this argument, the table of *a priori* probabilities shown below would be constructed.

Table 1
Table of probabilities for rolling two dice

x	$p(x)$	x	$p(x)$
2	$1/36$	8	$5/36$
3	$1/18$	9	$1/9$
4	$1/12$	10	$1/12$
5	$1/9$	11	$1/18$
6	$5/36$	12	$1/36$
7	$1/6$		

This $p(x)$ peaks at 7. MonopolyTM players and gamblers take note! The expected value of the experiment is given by

$$\begin{aligned} \langle x \rangle &= \sum p(x)x = 2/36 + 3/18 + 4/12 + 5/9 + 30/36 + 7/6 \\ &\quad + 40/36 + 9/9 + 10/12 + 11/18 + 12/36 = 7.00. \end{aligned}$$

Here, the most likely outcome and the expected outcome are the same. If you were to bet money on an outcome of the roll of dice, you could expect to lose money. If, for instance, you bet on 7, the chances are $1/6$ that it will occur. The chance that it won't is $5/6$.

Exercise: Try constructing the equivalent to the table above, if you have two 5 sided dice (There are such things!).

Exercise: If you have access to the spreadsheet program Excel[®], you can examine the behavior of a pair of dice without wearing out your desktop. Excel[®] has a feature under the "Options" box in the Toolbar called "Analysis Tools." If you now double click Analysis Tools, the first time there will be a pause as the add-on program is loaded. When the menu box comes up you will see a list of options, containing among other things, "Histogram" and "Random Number Generation." Double click Random Number Generation.

Notice that one box is labeled "Distribution." If you double click the arrow on the right of this box, a list will appear that contains among other things, "Normal" and "Discrete." Click Discrete. Now, click "Number of

Variables.” Here, type in 2. Now click “Number of Random Numbers” and insert the number of rolls you want. I suggest 100 for the first time. The prescription here should generate a 2 column by 100 row output. Finally, you need to tell the random number generator what range of numbers and their probabilities to use. You need to have made a column of the possible outcomes of the roll of a die, i.e. 1,2 3,... to 6. In a parallel column, type 1/6. 1/6,... For the last entry, type 1 — Sum of the other probabilities. My version of Excel[®] has a little glitch. It insists that the sum of the probabilities be 1, but its own arithmetic is not accurate enough to make the sum of the 6, 1/6s close enough to 1 for it. Click “Parameters” and insert the range where you have already inserted the properties, e.g. E1:F6. Now, click onto the “Output Range” box. In here you should insert the starting address of the column you wish to contain the random number output, e.g. A1. If you now click “OK,” Excel[®] will fill the indicated range with random integers, from 1 to 6, all with equal probability. You can now compute the sum of each pair of random numbers in the two output columns. This sum column simulates the sum of two dice.

You are now in a position to examine what the generator has wrought. First, estimate the expected value by using the Excel[®] function Average() on the original, un-summed output columns. Is the result 3.5000? Why not?

Next, find another short column in the spread sheet and put in the numbers 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12. I assume here that you used the “D” column, starting with 1. Now, go back into Options and click Analysis Tools. Select “Histogram.” Click “Input Range.” Insert the range containing the numbers you want to analyze, here C1:C100. Click “Bin Range.” Here you insert the histogram points, D1:D11. When you click “Output,” insert the range you want the output to appear in, say G1. Before you do anything else, click off the box “Pareto” and the box “Cumulative.” If those boxes are cleared, click “OK.” If your machine is set up like mine, you will see a chart flash by and then disappear. You should see the list of the number of random numbers in each range listed in a column, starting with G1. Depending on the version of Excel[®] you are using, the chart will either appear next to the histogram or on a separate sheet. If it is on a separate sheet, and you wish to see the graph, go up and click “Window.” Under this should be listed something called “Chart1.” Click this and you should then see your histogram.

Compare your measured result to Table 1. According to Table 1, we should expect $100/36$ points in the “2” bin and $100/18$ in the “3” bin. Typically, the results should not be exactly what was predicted. There should be more numbers in the “6” and “7” boxes than in the 2 box, but the result should not be exactly the prediction of Table 1. As with anything involving random numbers, there are variations in the output from what is expected. If Mother Nature wanted you to know things precisely, she never would have put them in the form of random numbers. But, as we shall see in this text, there are patterns and structures that we can find.

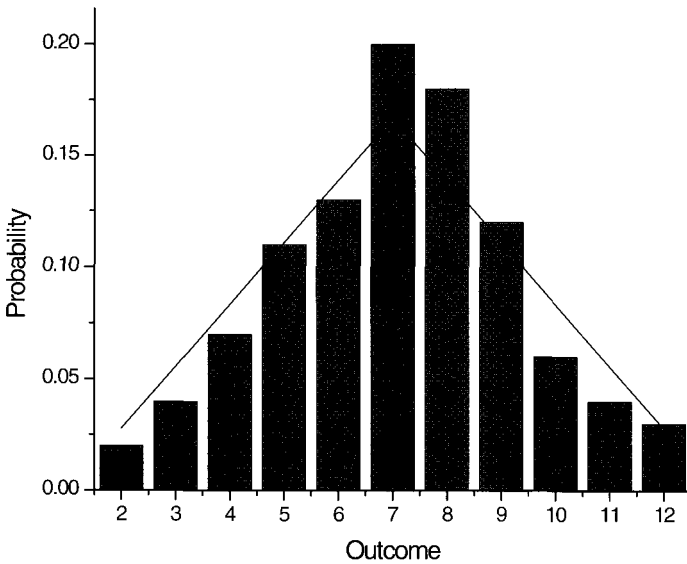


Fig. 1.2 Typical Result of Relative Frequency Test. The bars are the measurements and the line is the theory.

Now that you are sophisticated in Excel, create 2 sets of 1000 random integers and look again at the histogram for the sum. This should look a lot more like Table 1. In general, in measurement of random variables, more is better. More on this later. Save this spreadsheet and remember how you made it.

I should point out here that random numbers come in two flavors: discrete and continuous. An example of a system with a discrete outcome is the result of rolling a die. An example of a continuous random outcome would be the voltage sampled from a noisy electrical circuit. Typically, a noise voltage could be any number between -10v and 10v .

For a random variable with a continuous output, the probability for finding an output in the interval x , dx is defined as

$$p(x)dx. \quad (1.3.4)$$

The definition of the expected value for a continuous random variable that ranges from a to b is

$$\langle x \rangle = \int_a^b p(x)x dx. \quad (1.3.5)$$

Distinct random numbers also have a further division into finite outcomes and infinite possible outcomes. The number of counts measured in a set time for a radioactive decay experiment must be an integer, but in principle, could be any number between zero and infinity.

1.3.1 *Properties of probability density functions*

Probability distribution functions have some interesting and important properties:

$$\sum_{\text{all } x} p(x) = 1, \quad (1.3.6)$$

for discrete variables and

$$\int_{\text{all } x} p(x)dx = 1, \quad (1.3.7)$$

for continuous variables. In other words, the sum of the probabilities of all the possibilities is one.

For a discrete random variable, if $p(x)$ is the probability of an event happening, $1 - p(x)$ is the probability that it won't happen. For the continuous case,

$$\int_a^b p(x)dx \quad (1.3.8)$$

is the probability of an outcome between a and b ,

$$1 - \int_a^b p(x)dx \quad (1.3.9)$$

is the probability that the outcome will not be between a and b .

If $p(x_1)$ and $p(x_2)$ are the respective probabilities of two *independent* events x_1 and x_2 , the probability of x_1 and x_2 is

$$p(x_1)p(x_2).$$

If the probability of x_1 and x_2 is $p(x_1)p(x_2)$, the variables x_1 and x_2 are said to be *statistically independent* variables.

If

$$p_1(x_1) \text{ and } p_2(x_2) \quad (1.3.10)$$

are the respective probabilities of two mutually exclusive events x_1 and x_2 , the probability of x_1 or x_2 is

$$p_1(x_1) + p_2(x_2). \quad (1.3.11)$$

Note that 1.3.9 could have been derived from this property.

If $p_x(x)$ is the pdf for random variable x and $p_y(y)$ is the pdf for the independent random variable y , the pdf for the sum $z = x + y$ is given by

$$p(z) = \sum_x p_y(z - x)p_x(x) \quad (1.3.12)$$

for discrete variables and,

$$p(z) = \int_x p_y(z - x)p_x(x)dx \quad (1.3.13)$$

for continuous variables.

This property was used implicitly in the derivation of Table 1.

1.4 The Expected Value Process

You may compute the expected value of any function of x , $f(x)$ as follows,

$$\langle f(x) \rangle = \sum_x f(x)p(x), \quad (1.4.1)$$

for discrete variables.

In this case of a continuous variable, the expected value is computed by

$$\langle f(x) \rangle = \int_{\text{all } x} f(x)p(x) dx, \quad (1.4.2)$$

$$\langle x \rangle = \int_{\text{all } x} xp(x) dx. \quad (1.4.3)$$

An example of the use of such a distribution is shown below:

Let V be a fluctuating, random voltage applied to a circuit. The instantaneous power is V^2/R , where R is the resistance of the circuit. The expected average power is given by

$$\int \frac{p(V)V^2}{R} dV.$$

The process of computing the expected value of something is a linear operation; *i.e.*,

$$\langle \alpha f(x) + \beta g(x) \rangle = \alpha \langle f(x) \rangle + \beta \langle g(x) \rangle, \quad (1.4.4)$$

where α and β are fixed numbers and $f(x)$ and $g(x)$ are any functions of the random variable x . The proof is rather straightforward. Take the continuous case.

$$\begin{aligned} \langle \alpha f(x) + \beta g(x) \rangle &= \int_{\text{all } x} p(x)(\alpha f(x) + \beta g(x)) dx \\ &= \int_{\text{all } x} \alpha p(x)f(x) dx + \int_{\text{all } x} \beta p(x)g(x) dx \\ &= \alpha \int_{\text{all } x} p(x)f(x) dx + \beta \int_{\text{all } x} p(x)g(x) dx \\ &= \alpha \langle f(x) \rangle + \beta \langle g(x) \rangle. \end{aligned}$$

This property will be used often in later sections.

1.5 Variance and Standard Deviation

You often want to measure the spread of the likely outcomes about the mean. How wild a random variable is it? There are many ways to calculate a measure of the spread of a random variable, but it is usually done by computing the “variance”, σ^2 , of the probability distribution function.

$$\langle (x - \langle x \rangle)^2 \rangle \equiv \sigma^2 = \sum (x - \langle x \rangle)^2 p(x). \quad (1.5.1)$$

If we expand the formula for calculating the expected value of a function of x shown above, we can express the calculation of the variance in another way.

$$\begin{aligned}\sigma^2 &= \sum (x^2 - 2\langle x \rangle x + \langle x \rangle^2) p(x) \\ \sigma^2 &= \sum x^2 p(x) - 2\langle x \rangle \sum x p(x) + \langle x \rangle^2 \sum p(x) \\ \sigma^2 &= \langle x^2 \rangle - 2\langle x \rangle \langle x \rangle + \langle x \rangle^2 \\ \sigma^2 &= \langle x^2 \rangle - \langle x \rangle^2.\end{aligned}\tag{1.5.2}$$

The quantity $\langle x^2 \rangle$ is the expected value of x^2 . For the pair of dice,

$$\langle x^2 \rangle = 54.833, \quad \langle x \rangle^2 = 49$$

$$\sigma_x^2 = 5.833$$

$$\sigma_x = 2.415.$$

There are other measures for the deviation of a random variable from its expected value such as $\langle |x - \langle x \rangle| \rangle$, but the variance is usually the easiest to compute.

The square root of the variance, the *standard deviation*, of the probability function is the desired measure of the width of the function. This measure is particularly useful if the payoff from the outcome is a function of how much a given outcome differs from the expected value. Typically, if $\sigma/t\langle x \rangle$ is small, the distribution is said to be narrow. The variations about the expected value are not large compared to the expected.

These concepts can be extended to any experiment where the output has a random component.

Likewise, for a continuous variable,

$$\sigma^2 = \int (x - \langle x \rangle)^2 p(x) dx = \langle x^2 \rangle - \langle x \rangle^2.\tag{1.5.3}$$

1.6 Moments and Moment Generating Functions

The n th moment of a random variable is defined as $\langle x^n \rangle$. The n th central moment is defined as

$$\langle (x - \langle x \rangle)^n \rangle.$$

As we have seen, knowledge of the first moment and the second central moment (the variance) gives us some information about the shape of the pdf, but the description so far is incomplete. For instance, the mean and the variance give no clue as to whether the pdf is symmetrical about some point.

It is easy to show that, if the pdf is symmetrical about the expected value, the odd central moments are zero. Hence a measure of the asymmetry of a distribution can be given by the normalized third central moment, *viz.*,

$$S = \langle (x - \langle x \rangle)^3 \rangle / \sigma^3. \quad (1.6.1)$$

If this quantity is non-zero, the distribution is not symmetric. A positive value indicates a distribution with a longer positive tail than the negative tail. A negative value indicates the reverse.

A complete knowledge of the moments is equivalent to a complete knowledge of the pdf and vice versa. One way this can be seen is by considering the *Moment Generating Function*, or the *Characteristic Function* of the pdf, defined as follows:

$$Q_x(s) \equiv \int e^{-sx} p(x) dx = \langle e^{-sx} \rangle. \quad (1.6.2)$$

If the range of x is contained within $(0, \infty)$, s is usually taken as real and the integration is the Laplace transform of the pdf. In this case, Q_x is called the moment generating function.

Expand e^{-sx} in a Taylor's series and integrate,

$$\begin{aligned} Q_x(s) &= \int_0^\infty \left(1 - sx + \frac{s^2 x^2}{2} - \frac{s^3 x^3}{3 \times 2} + \dots \right) p(x) dx \\ &= \int 1 p(x) dx - s \int x p(x) dx + \frac{s^2}{2} \int x^2 p(x) dx - \frac{s^3}{3 \times 2} \int x^3 p(x) dx \dots \\ &= 1 - s \langle x \rangle + \frac{s^2 \langle x^2 \rangle}{2} - \frac{s^3 \langle x^3 \rangle}{3 \times 2} + \dots \end{aligned}$$

The n th moment is given by

$$\langle x^n \rangle = (-1)^n \left. \frac{d^n Q_x(s)}{ds^n} \right|_{s=0}. \quad (1.6.3)$$

If the range of x includes negative numbers, it is customary to keep s real, but to substitute is for s , $i = \sqrt{-1}$, in 1.6.2, and extend the integration from $(-\infty, \infty)$. In this case, you are doing a Fourier Transform and the resulting function is called the characteristic function. The formula for the moments is essentially the same, except that (i^n) is substituted for $(-1)^n$.

Similar calculations can be done for the central moments.

1.7 Common Types of Distributions

In this section I will describe some of the pdfs that often appear in the literature. They all have names. The pdf indicated above for the sum of the result of rolling three dice does not have a name, but is a well defined distribution. So please understand that the distributions described below are only a representative sample of the class of distributions.

1.7.1 Uniform distribution

This distribution is the simplest of all. It can describe both continuous and discrete outputs. In either case, the pdf is either a constant or zero. For the discrete case,

$$p(x_i) = \begin{cases} 1/N; & 1 \leq i \leq N \\ 0; & \text{otherwise} \end{cases}. \quad (1.7.1)$$

For the continuous case,

$$p(x) = \begin{cases} 1/(c-b); & b \leq x \leq c \\ 0; & \text{otherwise} \end{cases}. \quad (1.7.2)$$

Using either the moment generating function for the continuous uniform distribution or using the definition of the moments directly computed from the pdf, you would get

$$\langle x \rangle = \frac{c+b}{2} \quad \text{and} \quad \langle x^2 \rangle = \frac{b^2 + bc + c^2}{3}.$$

The variance is thus

$$\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2 = \frac{(c-b)^2}{12}.$$

There are uniform generators in Excel[®] and Mathematica[®] that seem to work well. Understand that random number generators on computers are all so-called pseudo-random generators. For example, if you start the generator out at the same value (called the random seed), the generator will always give the same output sequence. A good pseudo-random generator will not display an obvious correlation between the output values and will closely mimic the pdf of the distribution. Typically, a computer random generator will give you the choice of picking the random seed or it will generate its own, sometimes by manipulating a value obtained from the computer's clock.

1.7.2 Binomial distribution

Consider a situation where you have n attempts to shoot a foul shot in basketball. Let μ be the probability for each shot that you make the shot. The probability that you will miss the shot is $1-\mu$. We also assume that the probability for each shot is the same as all the others and independent. It should be immediately obvious that the chances that you will make none of the n shots is given by $p(0) = (1-\mu)^n$ (I know you think this is impossible, but I include it for mathematical completeness.)

The probability, $p(1)$, of making exactly one shot out of the total is given by $p(1) =$ (the probability of making the first shot) \times (probability of making no other shot) $+$ (the probability of making the second shot) \times (the probability of making no other shot) $+$...

$$p(1) = \underbrace{\mu(1-\mu)^{n-1} + \mu(1-\mu)^{n-1} + \dots}_{(n \text{ times})}$$

$$p(1) = n\mu(1-\mu)^{n-1}.$$

By a similar argument, you can show that $p(2) =$ (number of ways of making two shots in n tries without regard to order) times (The number of pairs in n) $\times \mu^2(1-\mu)^{n-2}$.

The symbol

$$\binom{n}{m} = \frac{n!}{(n-m)!m!}$$

denotes the number of ways of putting m things in n places without regard to order. It is called the Binomial Coefficient.

$$p(2) = \binom{n}{2} \mu^2 (1 - \mu)^{n-2}. \quad (1.7.3)$$

In general,

$$p(m) = \binom{n}{m} \mu^m (1 - \mu)^{n-m}. \quad (1.7.4)$$

This is the probability of making m shots in n tries if the probability for each shot is μ . It is called the *binomial distribution*.

The definition of the binomial distribution may seem a bit arcane and not really related to everyday life, but, in point of fact, the binomial distribution is by far the most often used distribution that affects your life. It is the distribution used to tell you how dangerous a pesticide might be. It is used to tell you how many people would vote for a given candidate. It is used to determine which TV shows are the most popular. A direct consequence of this is that the binomial distribution is often the most abused distribution. It is used to tell you that coffee is bad for you. Then, it is used in another study to tell you that coffee is good for you. . . Usually, the problem with clearly interpreting survey results is knowing whether the sample taken was really representative of the entire population you are extrapolating to. For instance, a telephone poll by definition does not sample people without telephones. It is conceivable that in some situations people without telephones will have a different response to a survey question than someone with a telephone. The reader may skip directly to the definitions of other distributions if you are not interested in the details of how pdfs like the binomial distribution help inform us about what happens in the real world.

A situation in which the binomial distribution is often used is when you know there are n events in an interval L . You wish to know the probability for m events in a sub-interval L_s . If the probability of an event is spread equally over the interval, $\mu = L_s/L$. The probability of m events in the interval is given by 1.7.4. The parameters n , m , and μ are used in exactly the same way.

Suppose you select 11 numbers from a continuous uniform random number population on the interval $[0,1]$. The probability of a number being in the interval, $[0,0.25]$ is $1/4$. What is the probability that only 2 of the 11 random numbers ended up in the interval $[0,0.25]$?

$$p(2) = \binom{11}{2} (1/4)^2 (3/4)^9 = 0.26.$$

The binomial distribution is most often encountered when you have a known number of independent trials of a phenomenon whose output can be one of two choices. For instance, if you do n trials where the outcome can be 1 or 2 and the probability of each 1 is μ , the probability of encountering m type one events in the trial is given by 1.7.4.

The binomial distribution is also appropriate when you have an experiment where n outcomes are distributed among k possibilities. If the probability for the p th event is μ_p , the probability of finding m_p events of the type p is also given by 1.7.4, *viz.*,

$$p(m_p) = \binom{n}{m_p} \mu_p^{m_p} (1 - \mu_p)^{n - m_p}. \quad (1.7.5)$$

This latter case is the use to which the binomial distribution will be most often used in this text. It explains the statistics for a measured pdf, a histogram. Since this is a text, there will be many examples given of measured pdfs and the results will never be exactly what is predicted to be the expected distribution. The statistics of the measured pdfs will be elucidated using the binomial distribution.

Exercise:

$$1 = (\mu + 1 - \mu)^N = \sum_{M=0}^N \binom{N}{M} \mu^M (1 - \mu)^{N-M}$$

Using the above definition, show that

$$\langle m \rangle = N\mu. \quad (1.7.6)$$

Hint:

$$\binom{N}{0} = 1.$$

The variance of a binomial distribution is

$$N\mu(1 - \mu). \quad (1.7.7)$$

You are doing measurements of a system with ten possible outcomes. The probability of the third outcome is 0.15. If you do 200 trials, what is the probability of getting 32 occurrences of the third outcome?

1.7.2.1 Using the binomial distribution

A popular use of the binomial distribution occurs when you are not so much looking at how many successes you have, but are trying to figure out from knowing the number of successes what the probability of the outcome is. Suppose you figure out how to select a set of N “typical” television viewers. If you find that a certain fraction m watch “Gilligan’s Island” reruns every week, how good an estimate is m/N of the probability for the entire population of television watchers? This scenario should be known to you by now. The company doing the measurement claims that they sampled about 1000 people and claim to be able to tell us the result for several hundred million television viewers. In a similar vein, doctors do epidemiological studies and tell us on alternate days that power lines cause brain cancer and then that they don’t. Again, how much should you believe them?

Figuring out the quality of survey results can be done by examining the amount of variation expected for the result compared to the result itself. If we get a probability of an outcome of 0.13 and the expected variation in the measurement, the standard deviation, is 0.27, we shouldn’t have a lot of confidence in the result. Using the results of the homework problems above, we can define a characteristic relative error as the expected standard deviation divided by the expected mean. By characteristic, I mean that it can generate numbers that tell you the order of magnitude of what you are dealing with. So, if the characteristic ratio of those quantities is 1%, then a variation of 10% would be considered very unusual. Likewise, if you calculate a characteristic ratio of 10%, then a 1% variation would be likewise rare.

According to the homework

$$\frac{\text{standard deviation}}{\text{mean}} = \frac{\sqrt{N\mu(1 - \mu)}}{N\mu} = \sqrt{\frac{(1 - \mu)}{N\mu}}.$$

If you are doing election polling, it is only worth it if $\mu \approx 0.5$, i.e. both candidates have a possibility to win. In that case the relative error is approximately $\sqrt{1/N}$. So, if you go out and get results from 1000 “typical” voters, the relative error should be about 3%. If you poll 10,000 voters instead, your relative error only goes down to 1%. Typically, national polls use on the order of 1000 samples. It tends not to be worth it to go the extra mile and use 10,000 or more samples.

If you are trying to estimate the percentage that are watching “Gilligan’s Island” daily, that should be less than 5%, so the relative error for a sample of 1000 people will be

$$\sqrt{0.95/0.05 * 1000} = \sqrt{0.19} = 0.14.$$

In this case you get a 14% error with 1000 samples. You expect to get 50 positive measurements, but the laws of chance say that you could easily measure as low as 43 measurements or a high as 57 measurements. You would still come up with the result that Gilligan’s Island is not very popular, but the result would not be as relatively accurate as for an event whose probability was near 0.5.

I will do one more example before moving on to numerical experimentation. Assume you are trying to discover whether a local plant’s effluent is causing some rare disease. In a situation like this, you know the background level of disease occurrence. This would be the random occurrence of the disease without a specific cause. By the definition of “rare” condition, the probability is small. For sake of argument, let us assume that $\mu = 0.001$. If you found 2 occurrences of the disease in 500 samples, do you have good enough evidence to go to court and stop the plant from operating?

Homework: Calculate the probability of measuring a 0.4% rate if the actual rate was 0.1%. Is the probability low enough to make it improbable that the conclusion that the plant was causing a problem was in error?

1.7.3 Poisson distribution

Consider a situation where a line of length L is divided into n segments by cuts of length Δx . Let the probability of one particle being in a line segment be given by,

$$p = \mu \Delta x. \tag{1.7.8}$$

Using the binomial distribution just derived, the probability of there being exactly m particles in the interval L is given by

$$\begin{aligned} p(m) &= \binom{n}{m} (\mu\Delta x)^m (1 - \mu\Delta x)^{n-m} \\ &= \binom{n}{m} (\mu L/n)^m (1 - \mu L/n)^{n-m}. \end{aligned} \quad (1.7.9)$$

If you let $\Delta x (= L/n)$ shrink to zero by letting n go to infinity while keeping the length constant, it is easy to show that

$$\lim_{n \rightarrow \infty} p(m) = \frac{(\mu L)^m}{m!} e^{-\mu L}. \quad (1.7.10)$$

This is the probability distribution for m events occurring in the interval, if the expected number of events is μL . It is called the *Poisson distribution*. This distribution occurs in many physical situations where the process has an output consisting of the sum of discrete, independent events. Photon counting experiments or radioactive experiments are typical situations where this distribution occurs. Recall that the binomial distribution is used when you know how many outcomes you have and you want to know how they probably are distributed. The Poisson distribution is used when you know the average number of outcomes and you want to know what the probability of a particular outcome is.

Computing the Poisson distribution as a limit of the Binomial distribution hints that in the limit of large n , we can approximate a Binomial Distribution with a Poisson distribution, *i.e.*

$$p(m) = \binom{n}{m} \mu^m (1 - \mu)^{n-m} \approx \frac{(\mu n)^m}{m!} \exp[-\mu n].$$

I often use this approximation when doing calculations on histograms. For example, calculate the probability of there being 10 counts in an interval for a uniform distribution divided into 100 bins and 1000 trials. The probability per bin is 0.01.

$$p(10) = \binom{1000}{10} (0.01)^{10} (1 - 0.01)^{990} \approx 0.1257.$$

But,

$$\frac{(10)^{10}}{10!} \exp[-10] \approx 0.1251.$$

Close enough.

Computation of the moments of this distribution is rather straightforward.

$$\langle m \rangle = \sum_{m=0}^{\infty} \frac{m(\mu L)^m}{m!} e^{-\mu L}.$$

To simplify the notation here, let $\mu L = \lambda$.

$$= e^{-\lambda} \left(\frac{d}{d\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} \right) \lambda.$$

But the sum term is the series for e^λ , thus

$$\langle m \rangle = e^{-\lambda} \lambda \frac{d}{d\lambda} e^\lambda = \lambda$$

$$\langle m \rangle = \lambda = \mu L. \quad (1.7.11)$$

$$\begin{aligned} \langle m^2 \rangle &= e^{-\lambda} \sum_{m=0}^{\infty} \frac{m^2 \lambda^m}{m!} \\ &= e^{-\lambda} \left(\lambda^2 \sum_m \frac{m(m-1)\lambda^{m-2}}{m!} + \lambda \sum_m \frac{m\lambda^{m-1}}{m!} \right) \\ &= e^{-\lambda} \left(\lambda^2 \frac{d^2}{d\lambda^2} e^\lambda + \lambda \frac{d}{d\lambda} e^\lambda \right) \end{aligned}$$

$$\langle m^2 \rangle = \lambda^2 + \lambda. \quad (1.7.12)$$

Thus,

$$\langle m^2 \rangle - \langle m \rangle^2 = \lambda. \quad (1.7.13)$$

The mean and the variance of a Poisson distribution are both λ .

Example:

The arrival of electrons at an anode, or photons from a coherent source can both be described by a Poisson process. If μ is the mean rate of arrival, the expected number measured in a time Δt is $\mu \Delta t$. The probability of n electrons (or photons) arriving in time Δt is given by

$$p(n) = \frac{e^{-\mu \Delta t} (\mu \Delta t)^n}{n!}.$$

The variance of the number of arrivals in Δt seconds is $\mu \Delta t$.

A measure of the quality of the measurements S/N would be the RMS fluctuation in the measurements divided by the mean.

$$S/N = \frac{\mu\Delta t}{\sqrt{\mu\Delta t}} = \sqrt{\mu\Delta t}.$$

The ratio of the RMS fluctuations to the mean varies as the square root of the mean. This effect can be the primary source of noise in an experiment involving weak light or small signals. The fluctuations in the measurement rate of photons coming even from a steady source appear as a noise that can tend to obscure the signal.

Excel[®] and Mathematica[®] contain good Poisson generators.

1.7.4 Gaussian (or normal) distribution

This distribution is defined for continuous variables.

$$p(x)dx = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx. \quad (1.7.14)$$

The expected value of x is μ and the variance of x from the mean is σ^2 , i.e.,

$$\langle (x-\mu)^2 \rangle = \sigma^2.$$

The other central moments of the Gaussian can easily be calculated using the characteristic function.

$$Q(s) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{isx} e^{-\frac{x^2}{2\sigma^2}} dx.$$

$$Q(s) = e^{-s^2\sigma^2/2}.$$

In particular, the fourth central moment will be of use later in this text.

$$\langle \Delta x^4 \rangle = (i^4) \left. \frac{\partial^4 Q(s)}{\partial s^4} \right|_{s=0} = 3(\sigma^2)^2. \quad (1.7.15)$$

Likewise, it can be shown that all the odd central moments are zero.

The Gaussian distribution is very useful in light of a theorem known as the *Central Limit Theorem*. It is stated as follows: If x is the average of n independent random variables from a population of expected value μ and

a standard deviation σ , in the limit as n goes to ∞ , the probability density for the variable

$$z = \frac{(x - \mu)}{\sqrt{2\sigma^2/n}} \quad (1.7.16)$$

becomes Gaussian, *i.e.*,

$$\lim_{n \rightarrow \infty} p(z) dz = \frac{1}{\sqrt{\pi}} e^{-z^2} dz. \quad (1.7.17)$$

In other words, as n gets large,

$$p(x) \rightarrow \frac{1}{\sqrt{2\pi\sigma^2/n}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2/n}\right). \quad (1.7.18)$$

There is a lot here. First, it says that if you average the samples of a random variable together, *the probability density of the resultant variable tends to a Gaussian no matter what the original probability distribution function.* Further, in the limit, the expected value of the average is the same as the expected value for the original population. Moreover, this formula claims that in the limit, the variance of the mean is less than the variance of the original population, going down by a factor of n , the number of points used in the average.

Example: Consider a system measuring the number of photons in a regular time interval such that the mean number per interval is 10. The probability distribution is Poisson so that the variance is 10. Figure 1.3 shows the pdf for a Gaussian of mean and variance of 10.0 (solid line), along with the Poisson pdf (shaded rectangles).

Recall that the distribution functions are often used as parts of sums or integrals, so that exact agreement at every point is not always needed for a useful approximation. As a rule, a Gaussian is a good approximation to a Poisson distribution if the expected value is greater than 10.

1.7.5 Student-*t* distribution

A useful distribution for statistical experiments is the Student-*t* distribution. It is used on the occasions where you have measured means and measured variances which are, by definition, estimates of the expected values.

$$f(t, \nu) = \frac{1}{\sqrt{\pi\nu}} \frac{\Gamma((\nu + 1)/2)}{\Gamma(\nu/2)} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2} \quad (1.7.19)$$

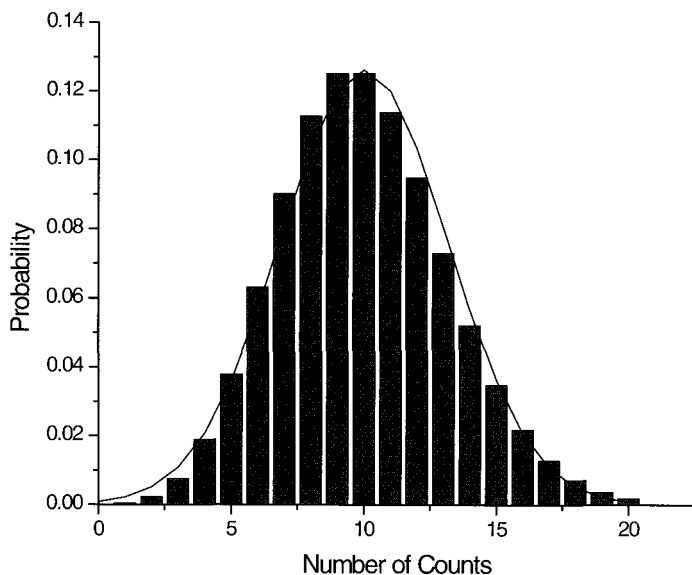


Fig. 1.3 Pdfs for a Gaussian (continuous line) and a Poisson (bars) both with a mean and variance of 10.

Here, t is the independent variable and ν is known as *the number of degrees of freedom*. Don't panic. Tables of the t -distribution are ubiquitous and there are functions in Excel[®] and Mathematica[®].

Typically, the t is made up as follows.

$$t = \frac{\bar{x}}{s},$$

where \bar{x} is the measured mean obtained using N points, and s is the measured standard deviation. The number of degrees of freedom is one less than the number of points used to estimate the variance, here $N - 1$. In principle, the t distribution applies only if the original random numbers had a Gaussian distribution. However, thanks to the central limit theorem, the distribution for the averages approaches a Gaussian distribution as more and more points are used for the average. In most statistics texts, the implicit assumption is that there are enough points used that the means derived have a Gaussian distribution.

1.7.6 Sum of two Gaussian variables

Let z be the sum of two independent Gaussian random variables with the same mean and variance, μ and σ^2 , respectively. From 1.3.13, the pdf for z is given by

$$\begin{aligned} & \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \exp\left(-\frac{(z-x-\mu)^2}{2\sigma^2}\right) dx \\ &= \frac{1}{\sqrt{4\pi\sigma^2}} \exp\left(-\frac{(z-2\mu)^2}{4\sigma^2}\right). \end{aligned} \quad (1.7.20)$$

1.7.7 The Chi-squared distribution

Let z_n be a zero mean Gaussian random variable with variance one. Then define

$$\chi_\nu^2 = \sum_{n=1}^{\nu} z_n^2.$$

The pdf for this variable is

$$p(\chi_\nu^2) = \frac{1}{(2^{\frac{\nu}{2}})\Gamma(\frac{\nu}{2})} (\chi_\nu^2)^{\frac{\nu-2}{2}} \exp\left[-\frac{\chi_\nu^2}{2}\right]. \quad (1.7.21)$$

It can be shown that

$$\langle \chi_\nu^2 \rangle = \nu \quad (1.7.22)$$

and

$$\sigma_{\chi^2}^2 = 2\nu. \quad (1.7.23)$$

This distribution is most useful in dealing with the statistics of estimates of the variance. Let $z_n = (x_n - \bar{x})/\sigma$. Then,

$$\chi_{\nu-1}^2 = \sum_{n=1}^{\nu} z_n^2 = \sum_{n=1}^{\nu} \frac{(x_n - \bar{x})^2}{\sigma^2} = (\nu - 1) \frac{s^2}{\sigma^2}.$$

Note the $\nu - 1$ that comes about because you have to use the computed average to estimate the variance.

In the limit of large ν , the distribution becomes Gaussian and the chi-squared distribution can be written,

$$p(\chi_\nu^2)d\chi_\nu^2 = \frac{1}{\sqrt{4\pi\nu}} \text{Exp} \left[-\frac{(\chi_\nu^2 - \nu)^2}{4\nu} \right] d\chi_\nu^2.$$

It is straightforward to convert this to the pdf for the measured variance in the large number limit,

$$p(s^2)ds^2 = \frac{1}{\sqrt{4\pi(\nu-1)}} \text{Exp} \left[-\frac{\left(\frac{s^2}{\sigma^2}(\nu-1) - (\nu-1) \right)^2}{4(\nu-1)} \right] \frac{ds^2}{\sigma^2}(\nu-1)$$

$$p(s^2)ds^2 = \frac{1}{\sqrt{4\pi(\sigma^2)^2/(\nu-1)}} \text{Exp} \left[-\frac{(s^2 - \sigma^2)^2}{4(\sigma^2)^2/(\nu-1)} \right] ds^2. \quad (1.7.24)$$

1.7.8 The error function

The indefinite integral of a Gaussian is a very important quantity in statistics.

$$\frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt = \text{erf } x, \quad (1.7.25)$$

the *error function*.

In terms of the distributions commonly used,

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x \exp \left(-\frac{(t - x_0')^2}{2\sigma^2} \right) dt = \frac{1}{2} \left(1 + \text{erf} \left[\frac{(x - x_0)}{\sqrt{2}\sigma} \right] \right), \quad (1.7.26)$$

where it is understood that $\text{erf}(-x) = -\text{erf}(x)$.

Values for the error function may be obtained from tables found in most statistics texts or from functions found in Excel[®] or Mathematica[®].

There are many other examples of probability distribution functions, but those outlined above will be sufficient for this work. In any event, please do not adopt the standard procedure of assuming that every distribution is Gaussian!

1.8 Functions of More Than One Random Variable

Until now, we have restricted our discussion to situations involving just one random variable. It often occurs that we have to deal with functions of more than one random variable, where the random variables involved are not necessarily independent. An example would be the probability of a day in Cleveland above 90°F *and* there being a thunderstorm on that day. The two events are linked, but they are both random variables.

The probability of 1 *and* 2 is denoted $p(1, 2)$. The *conditional probability* of 1 *if* 2 occurs is denoted $p(1|2)$. That would be the probability of a thunderstorm *if* the temperature were above 90°F. There are several useful theorems relating these probabilities:

1.8.1 (*Bayes theorem*)

$$p(1, 2) = p(1|2)p(2) = p(2|1)p(1). \quad (1.8.1)$$

The probability of 1 and 2 is the conditional probability of 1 given 2 times the probability of 2. In addition

$$p(1) = \sum_2 p(1, 2). \quad (1.8.2)$$

The probability distribution of the outcome $p(1)$ is obtained if you sum $p(1, 2)$ over all 2 type events. For continuous random variables, you have

$$p(x_1) = \int p(x_1, x_2) dx_2. \quad (1.8.3)$$

These results generalize for a set of more than two random variables, *i.e.*,

$$p(x_1, \dots, x_k) = \int p(x_1 \dots x_n) dx_{k+1} \dots dx_n. \quad (1.8.4)$$

Characteristic functions can be defined for joint pdfs:

$$Q_{xy}(s_1, s_2) = \int e^{-(s_1 x_1 + s_2 x_2)} p(x_1, x_2) dx_1 dx_2. \quad (1.8.5)$$

These concepts are most useful when considering functions of multiple random variables. In general the expected value of a function of multiple

random variables is given by

$$\langle f \rangle = \iiint \cdots \int f(x_1, x_2, x_3, \dots, x_k) p(x_1, x_2, x_3, \dots, x_k) dx_1 dx_2 dx_3, \dots, dx_k. \quad (1.8.6)$$

This formula is also useful for computing multiple moments.

Define

$$\Delta x_i \equiv x_i - \langle x_i \rangle. \quad (1.8.7)$$

Then

$$\langle \Delta x_1 \Delta x_2 \rangle = \iint \Delta x_1 \Delta x_2 p(x_1, x_2) dx_1 dx_2. \quad (1.8.8)$$

This quantity is known as the *covariance* of x_1 and x_2 . It is a measure of the correlation in the fluctuations of the two variables from their means.

An alternate method of calculating the moments is through the moment generating function of the characteristic function. The multiple moments of $p(x)$ are given by

$$\langle x_1 x_2 \dots x_k \rangle = [-1]^k \frac{\partial^k Q(\mathbf{s})}{\partial s_1 \partial s_2 \dots \partial s_k}. \quad (1.8.9)$$

Consider a simple example where $x_3 = x_1 + x_2$, where x_1 and x_2 are both random variables. We wish to know the pdf for the new random variable x_3 . If we could find the pdf for x_3 and x_1 , the problem could be handled rather simply. By 1.8.3,

$$p(x_3) = \int p(x_3, x_1) dx_1.$$

It can be shown that, if x_3 is a function of x_1 and x_2 ,

$$p_{12}(x_1, x_2) dx_1 dx_2 = \left| \frac{\partial x_3}{\partial x_1} \right|^{-1} p_{13}(x_3, x_2) dx_3 dx_2. \quad (1.8.10)$$

Here,

$$p_{12}(x_1, x_2) = p(x_3 - x_2, x_2)$$

and

$$\left| \frac{\partial x_3}{\partial x_1} \right| = 1,$$

so

$$p(x_3) = \int p_{12}(x_3 - x_2, x_2) dx_2.$$

If x_1 and x_2 are independent random variables, then

$$p(x_3) = \int p_1(x_3 - x_2)p_2(x_2)dx_2, \quad (1.8.11)$$

the convolution of the two pdfs.

Example: Let x_1 and x_2 have a uniform distribution from 0 to 1.

Then

$$p(x_3) = \int_0^{x_3} dx_2; \quad x_3 \leq 1; \quad p(x_3) = \int_{x_3-1}^1 dx_2; \quad x_3 > 1.$$

$$p(x_3) = x_3, \quad 0 < x_3 \leq 1, \quad p(x_3) = 2 - x_3, \quad 1 < x_3 < 2; \quad p(x_3) = 0$$

otherwise.

This result is equivalent to summing over the probabilities of all combinations of x_1 and x_2 that add up to x_3 .

Consider now computing the probability of $x_3 = x_1/x_2$, from the known probability of x_1 and x_2 , $p_{12}(x_1, x_2)$. We wish to make a variable change from (x_1, x_2) to (x_3, x_2) . For this problem,

$$p_{12}(x_1, x_2) = p_{12}(x_3x_2, x_2)$$

and

$$\left| \frac{\partial x_3}{\partial x_1} \right| = \left| \frac{1}{x_2} \right|.$$

Thus,

$$p(x_3) = \int p_{12}(x_3x_2, x_2) |x_2| dx_2.$$

Example: Again let x_1 and x_2 come from a uniform distribution from 0 to 1. For $x_3 \leq 1$,

$$p(x_3) = \int_0^1 x_2 dx_2 = 1/2.$$

For $x_3 > 1$,

$$p(x_3) = \int_0^{1/x_3} x_2 dx_2 = \frac{1}{2x_3^2}.$$

Similar forms apply for the computations involving discrete event pdfs. For instance, the function in Table 1 for the probability of the sum of the results of rolling two dice can be computed from the individual pdf for each die. Taking into account that each die is independent, we get

$$p(\text{sum}) = \sum_{x_1+x_2=\text{sum}} p_1(\text{sum} - x_2)p_2(x_2).$$

But we know that each p_i has the same value, $1/6$. Therefore, the calculation becomes

$$p(\text{sum}) = \frac{1}{36} \times (\text{number of terms where } x_1 + x_2 = \text{sum}).$$

Try to compute the pdf for the sum of three dice.

1.8.2 Joint Gaussian distributions

If you have multiple zero mean random variables that have a joint Gaussian distribution, the pdf for the data set would be written

$$p(x_1, x_2, \dots, x_n) = (2\pi)^{-n/2} |\mathbf{\Lambda}|^{-1/2} \exp(-0.5(\mathbf{x}) \cdot \mathbf{\Lambda}^{-1} \cdot \mathbf{x}), \quad (1.8.12)$$

where \mathbf{x} is the vector of random variables,

$$\mathbf{x} = (x_1, x_2, x_3, \dots), \Lambda_{ij} = \langle x_i x_j \rangle,$$

and $|\mathbf{\Lambda}|$ is the determinant of $\mathbf{\Lambda}$.

The multiple moments of this distribution can be calculated through use of the characteristic function.

$$Q(s_1, s_2, \dots, s_k) = (2\pi)^{-n/2} |\mathbf{\Lambda}|^{-1/2} \int_{-\infty}^{\infty} \exp(-0.5(\mathbf{x} \cdot \mathbf{\Lambda}^{-1} \cdot \mathbf{x})) \\ \times \exp(is \cdot \mathbf{x}) d\mathbf{x}, \quad (1.8.13)$$

By completing the square of the exponent of the integrand, we get

$$\begin{aligned} Q(s_1, s_2, \dots, s_k) &= (2\pi)^{-n/2} |\Lambda|^{-1/2} \\ &\quad \times \int_{-\infty}^{\infty} \exp(-0.5(\mathbf{x} - i\Lambda \cdot \mathbf{s}) \cdot \Lambda^{-1} \cdot (\mathbf{x} - \Lambda \cdot \mathbf{s})) \\ &\quad \times \exp(-0.5\mathbf{s} \cdot \Lambda \cdot \mathbf{s}) d\mathbf{x}. \end{aligned}$$

Now, performing the multiple integrations, we get

$$Q(s_1, s_2, \dots, s_k) = \exp(-0.5\mathbf{s} \cdot \Lambda \cdot \mathbf{s}). \quad (1.8.14)$$

The fourth multiple moment will later be of interest.

$$\begin{aligned} \langle x_1 x_2 x_3 x_4 \rangle &= \left. \frac{\partial^4 Q}{\partial s_1 \partial s_2 \partial s_3 \partial s_4} \right|_{s_1=s_2=s_3=s_4=0} \\ &= \Lambda_{12}\Lambda_{34} + \Lambda_{13}\Lambda_{24} + \Lambda_{14}\Lambda_{23}. \end{aligned}$$

Thus, the fourth central moment can be written

$$\begin{aligned} \langle x_1 x_2 x_3 x_4 \rangle &= \langle x_1 x_2 \rangle \langle x_3 x_4 \rangle + \langle x_1 x_3 \rangle \langle x_2 x_4 \rangle \\ &\quad + \langle x_1 x_4 \rangle \langle x_2 x_3 \rangle. \end{aligned} \quad (1.8.15)$$

If the variables are not zero mean, it can be shown that the characteristic function is given by

$$Q(s_1, s_2, \dots, s_k) = \exp(i\mathbf{s} \cdot \boldsymbol{\mu}) \exp(-0.5\mathbf{s} \cdot \Lambda \cdot \mathbf{s}), \quad (1.8.16)$$

where $\boldsymbol{\mu} = (\mu, \mu, \mu \dots)$, $\mu = \langle x_i \rangle$.

In this case

$$\begin{aligned} \langle x_1 x_2 x_3 x_4 \rangle &= \langle x_1 x_2 \rangle \langle x_3 x_4 \rangle + \langle x_1 x_3 \rangle \langle x_2 x_4 \rangle \\ &\quad + \langle x_1 x_4 \rangle \langle x_2 x_3 \rangle - 2\mu^4. \end{aligned} \quad (1.8.17)$$

1.9 Change of Variable

This section is devoted to examining, in more detail than before, the concept of computing the probability density function of a function of a random variable x if we know the probability density function for x .

Let $y(x)$ be a monotone (either increasing or decreasing) function of x . See Figure 1.4. From this figure, it can be seen that the probability of the function shown being in the interval $y(x)$, $y(x) + dy$ is the same as the

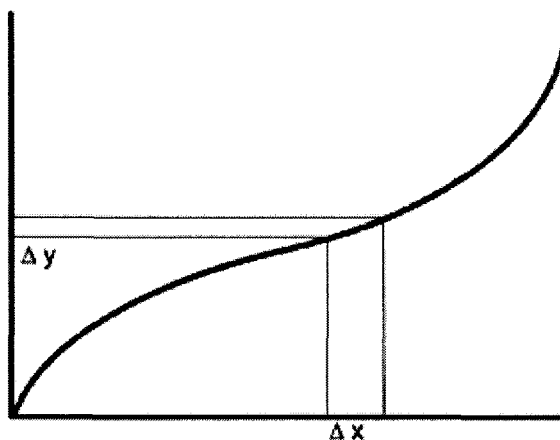


Fig. 1.4 Monotone function.

probability of x being in the interval $x, x + dx$. Further, dy is related to dx by

$$dy = \left| \frac{dy}{dx} \right| dx. \quad (1.9.1)$$

The absolute value sign comes about because we are only interested in the length of dy , not its sign. It follows then that the pdf for $y(x)$ is given by

$$p(y) = \frac{p_x(x(y))}{\left| \frac{dy}{dx} \right|}, \quad (1.9.2)$$

where the pdf on the right is the pdf for x .

Example:

Suppose x is a Gaussian random variable with mean μ . What is the pdf for $\ln x$?

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

$$g(x) = \ln x.$$

$$\frac{dg}{dx} = \frac{1}{x} = e^{-g} \cdot x > 0.$$

Thus,

$$p(g) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(g - \frac{(e^g - \mu)^2}{2\sigma^2}\right). \quad (1.9.3)$$

The pdf for the log is hardly Gaussian! This is known as a *log normal distribution*.

The formula for a change of variable is a bit more complicated if $g(x)$ is not a monotone function. See Figure 1.5.

If $g(x)$ lies within the interval $y, y + \Delta y$, then the corresponding x interval can lie at x_1 or at x_2 , or at x_3 , etc. We must take into account all the possible x intervals. Let $\{x_n\}$ be the set of solutions to $g(x_n) = y$.

$$p(y) = \sum_n \frac{p(x_n)}{\left|\frac{dg(x_n)}{dx}\right|}, \quad (1.9.4)$$

where the sum runs over all x_n .

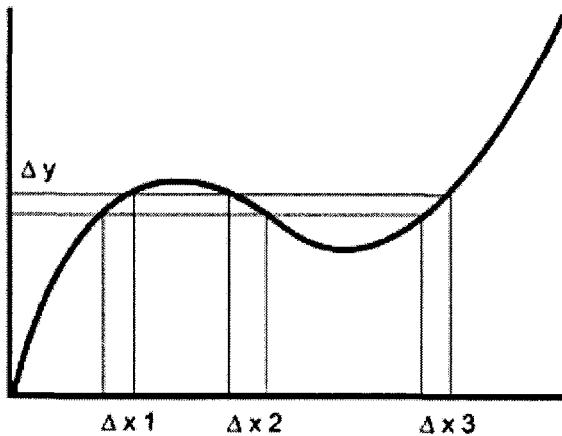


Fig. 1.5 Function with multiple roots.

Example:

Let $y = x^2$. For any y , there are two solutions $\pm x$. If the pdf for x is a Gaussian with a zero mean, the pdf for x^2 is derived as follows:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

$$g(x) = x^2; \frac{dg}{dx} = 2x = 2\sqrt{g}.$$

Thus,

$$p(g) = \frac{2}{2\sqrt{2\pi\sigma^2g}} \exp\left(-\frac{g}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2g}} \exp\left(-\frac{g}{2\sigma^2}\right). \quad (1.9.5)$$

Suggested Reading

1. "Statistical Design and Analysis of Engineering Experiments", Charles Lipson and Narendra J. Sheth, McGraw-Hill Book Company (New York), 1973.
2. "Probability and Statistics for Engineering and the Sciences", Second Edition, Jay L. Devore, Brooks/Cole Publishing Company, 1987.
3. "Facts from Figures", M. J. Moroney, Penguin Books, 1967.
4. "Statistical Signal Processing — Detection, Estimation, and Time Series Analysis", Louis L. Sharf, Addison-Wesley, 1991.